DOCUMENT RESUME

ED 119 610 IR 003 059

AUTHOR Geis, George L.
TITLE Trick or Treatment.
INSTITUTION McGill Univ., Montreal (Quebec). Center for Learning and Development.
PUB DATE 76
NOTE 7p.
JOURNAL CIT Learning and Development; v7 n3 Dec-Jan 1976

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Comparative Analysis; *Educational Methods; Educational Objectives; *Educational Research; Educational Researchers; Educational Strategies; Effective Teaching; Instructional Design; Methods Research; *Research Methodology; *Teaching Methods; Teaching Techniques

ABSTRACT

Comparative analyses of educational methods have often been inconclusive; there are several possible explanations: (1) each method is optimally effective under specifically differing circumstances; (2) the studies tend to emphasize one input variable to the exclusion of the others; (3) the methods do not have widely accepted definitions; and (4) the methods are often tested before they are perfected. Comparative studies could be enhanced by more rigorously defining the methods that are being investigated and by standardizing the laboratory techniques and observation schemes that are employed. In addition, the studies could be made more usable to the layman by spelling out what standards of educational effectiveness the researcher has employed. To obtain an answer for a specific comparison in a designated milieu, carefully controlled studies can be mounted using similar populations aimed at similar instructional goals and exposed to two differing treatments. (EMH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

learning and development

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

George L. Geis
CLD - McGill Univ.

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

Centre for Learning and Development
McGILL UNIVERSITY

"TRICK OR TREATMENT"

by
George L. Geis

Introduction

The recent decade or so might be called (among other names) the Age of Consumerism. Attention has been turned to the purchaser or user of services and products. It is becoming clear that being a discriminating consumer is not an easy job. Even the first step--formulating the most useful consumer questions--requires sophistication and training. The simple, straightforward questions often overlay unrecognized complex problem areas.

What do these consumer questions have in common:

The patient asks his dentist: "Is it better to fill a decayed tooth or extract it?"

The client asks the lawyer: "Is it better to take a case to trial or to settle out of court?"

The parent asks the child psychologist: "Is reward more effective than punishment?"

One common feature is that they all are likely to elicit the same answer from the expert:

"It all depends . . ."

The educational researcher faces a similar problem when he is asked (or he asks himself) which instructional technique ("method" or "treatment") is better. For example, "Is computer-assisted instruction superior to the traditional lecture method?" Consumers of education as well as researchers have for years asked this comparative question about a great variety of teaching methods. This paper attempts to analyze the question, to suggest when it is useful and when not, and to show how it must be reformulated to serve some purposes.

© CENTRE FOR LEARNING AND DEVELOPMENT

VOL. 7 NO. 3 December-January 1976

The comparative study, that is research that compares one teaching method with another, has had a long history. Extensive summaries of this methodological literature appear periodically. Some refer to a particular method (e.g., "Learning in Discussions: A resumé of the Authoritarian--Democratic studies."[1]) Some refer to comparisons of any alternative treatment* with the so-called traditional methodology (e.g., The Teaching-Learning Paradox: A Comparative Analysis of College Teaching Methods)[2]. In every case the ambiguity of results described by these vast surveys suggests the futility of continuing to search for a simple answer at this point, or at least suggests deferring additional research of this sort until there is sufficient clarification of the question.

The usual conclusion reached upon completion of one of these surveys is: there is no clear answer; no clear evidence of superiority; no real basis for choosing one method or another. Furthermore these reviews present a similar research history for each innovative method studied. (1) The originators of a new method show its dramatically greater effectiveness; (2) follow-up studies, some by acolytes of the innovators, begin to cast doubt on those first findings; (3) later studies often fail to show differences between treatments. And so in answer to the simple comparison question the expert reviewer is driven to emit that irritating phrase: "It depends . . ."

The irritation may be reduced if we understand some of the reasons for the failure of research in this area including the self-defeating nature of the question. Furthermore as we come to understand those reasons we can develop better research strategies for getting at some of the questions comparative studies aim to answer.

Educational researchers Richard C. Anderson and Gerald Faust[3] comment:

> In the abstract and without qualifications, it is very difficult to prove that one method of instruction is better than another. About all a single experiment can show is that one particular lesson (or set of lessons) is more effective than a comparison lesson. Suppose that a hypothetical experiment shows that students learn more from a series of lectures than from several textbook chapters covering the same material. Obviously one should not conclude from this study that all lectures are better than all textbooks. Nor should one conclude that the difference in achievement is due to superiority of the spoken word over the printed word. It may have been the case, for example that the lecturer expressed concepts more clearly or used more informative examples than the textbook, or that the percentage of students that attended class and listened to the lectures was greater than the percentage that actually read the textbook. (page 4)
>
> . . . There are many explanations for the inconclusive results of studies of teaching methods . . . first there is the obvious point that different methods may work best with different students, different purposes, and under different conditions.
>
> Second, there has been confusion about what is meant by a given method. Various investigators, for example, have defined methods differently. Naturally, variations in method could lead to variable results with students . . . (page 5)
>
> . . . Third, each method emphasizes a few processes and features of instruction but ignores others. In most 'methods studies' important factors have been overlooked.
>
> Finally--and, we believe, most important in the majority of the studies--the teaching methods

* We will use "treatment" to refer broadly to method, strategy or technique of instruction.

> under consideration have not been very good. Few investigators have taken pains to develop a method until it was demonstrably effective. Frequently, comparisions among methods have been premature. . ." (page 5)

Let us examine closely two problems implied in the above comments. Basically they are problems of definition. They are concerned with the referents in the sentence "Is Treatment A better than Treatment B? Specifically, the key problems--important contributors of the failure of comparative research--are the definitions of the methods employed (technically: the experimental independent variable) and of the word better (technically: the experimental dependent variable).

Defining the Method.

Very often an educational method is described solely in terms of its formal properties, for example, Computer Assisted Instruction. Apparently any instruction which is assisted by a computer qualifies as CAI. But surely specialists working in this area have other, more critical, characteristics in mind when they say CAI, for example: self-pacing and immediate feedback during instruction.

We would not think of asking a physician either "how effective are pills?" or "Are pills more effective than injections?" because we intuitively know that the formal description of the medication--the description in terms of how it looks--is almost irrelevant. Important are such things as what is in the pill and what is wrong with us when we take the pill.

This is a key problem that vexes the educational researcher who is asked a question about the comparative effectiveness of methods since the methods are usually inadequately defined. Identifying obvious physical characteristics (e.g., a pill) is not sufficient. Many things which share those characteristics may have quite different effects.

The referent problem becomes more obscured in some cases than others. At least with CAI the defining characteristic (i.e., the use of a computer) is objective. When in doubt the observer could call the Computing Centre and ask someone whether the object in question may properly be called a computer. But take the case of programmed instruction, or, lecture-discussion, or, modularized instruction. It is difficult to obtain agreement on whether or not a given piece of instructional material is or is not a program or a module, whether a class session was or was not really a lecture instead of a lecture-discussion.

Before one can determine whether one instance (a text) is a good exemplar of a class (e.g., programmed instruction) he or she should be able to state the common attributes of members of the class. Disagreement among observers at this point almost precludes further exploration of the original problem (i.e., "Does it work?")

Conversely, definitions, especially those based on formal properties may exclude instances that ought to be included. Methods of instruction that seem to be quite different on the face of it may share critical functional attributes. Looking only at the outside, the package, as it were, is likely to be highly misleading.

The standardization problem.

A pill may not be like another pill, except in terms of trivial, external attributes. But one aspirin tablet, at least most of the pharmaceutical advertisements tell us, is quite like another. There is a good degree of standardization in drugs as well as in many other product areas. Here our medical analogy breaks down. Instructional "products" differ widely even when called by the same name and even when they share many common characteristics.

Even though there may be agreement on the basis of objectively determined attributes that two instances are of the same class, there still are major differences between the two. Faced with two introductory text books both on the same subject, both of about the same length and aimed

at the same reading level, both with study questions, do-it-yourself experiments etc., the professor still is likely to label one as the better text. One speaker will seem superior to another lecturer in an invited lecture series.

The general questions "Does it work?" and "Is it better?" are unanswerable given the great differences between any two its one might choose. To put it another way: Generalizations about the effectiveness of any "method" must be severely limited. Since there are likely to be large differences among the instances of the same method, the results of a study involving one instance or even a few instances probably will not apply universally to all instances of that method.

It should be clear that comparisons between one instance of one method and one instance of another (e.g., my lecture versus your module) are especially likely to be specious.

The problem of definition becomes transparent, does it not, when one term of the comparison, for example, the traditional method, is so poorly defined that two examples of the traditional method could be used to re-form the question? Thus, the question: is lecture better than CAI? becomes almost meaningless if we know that professor A's traditional lecture is better than professor B's traditional lecture and such is usually the case.

If we can agree that within the term lecture, or programmed text, or film one could find the full range from very effective to totally ineffective examples in terms of teaching effectiveness, then obviously in a comparison between this method and some other much more explicit definition of the term is needed. Otherwise it is likely that in one study Lecture A will prove to be superior to CAI Lesson Type 1. While in a second study CAI Lesson Type 2 will produce better results than Lecture B. And that kind of paradox seems to permeate the method comparison literature, as we pointed out earlier.

Summary

The problems raised are not trivial ones or academic nit-picking. Almost every researcher who has reviewed a set of comparison studies echoes the words of Robert Hohn.[4]

> Inadequate description of the experimental techniques as well as control conditions, is perhaps the greatest deficiency in the recent literature on teaching innovation. A large majority of the thirty-one studies reviewed for this paper which compared more than one strategy of teaching provided incomplete information about both treatments employed. The typical procedure is to characterize a particular treatment with a label such as "lecture", "traditional", "self paced" or "group" with little or no data or operational terms used to clarify what particular interaction was occurring within these groups. (page 3)

The criterion problem.

Now let us look at the other half of the original questions, phrases like: "is better than" or "works well." What precisely does better mean?

Success (i.e., it works!) can refer to as many different criteria as there are observers. An instructional technique might be viewed as successful if it produces the designated changes in students for less cost, or in less time, or while simultaneously producing more favourable student attitudes toward the instruction. It may be judged as successful if it allows previously excluded groups to succeed in the instructional system, or provides enrichment for those students who are moving ahead rapidly. Success may mean that the method "impresses" students, faculty, government, alumni, etc. if it presents the image of progress (even though no evidence may have been gathered about its instructional effectiveness).

When we are deciding on the degree of success, do we plan to look merely at

one criterion (cost, teaching effectiveness) or at a wide set of outcomes?

Suppose we choose to look at effectiveness in terms of student achievement. In most studies all that is reported are gross summaries of achievement test scores. What do these "class quizzes" or "final grades" represent? They mean little if we do not know what the objectives of the instructional treatments were. One thing we do know is that it is very likely that different instructional objectives may best be reached by using different "methods." (A leading educational psychologist, Robert Gagné has recently devoted an entire book to exploring this very topic, contending throughout that there are optional methods for each general type of educational objective.)[5] For example it is likely that the best way to learn a large number of simple associations (e.g., the names of the bones of the foot) is not the best way to learn high level decision making skills (e.g., determining at which point in a river the addition of a dam would have the fewest negative effects on the ecological system).

In developing criteria for success one is faced with making explicit the general and specific purposes of the evaluation. If the decision to be made about a new innovation is going to be primarily based on problems of funding, then, obviously, data on cost ought to be gathered. If the observer is interested in the cognitive impact of an instructional method, he would not attach too much weight to the student attitude data supplied by the innovator. If the observer is interested in higher level cognitive objectives, results from class tests on fact-recall should not impress him.

Are comparison studies valid?

We have just stated that evaluation is purposeful. The question: "Is Treatment A superior to Treatment B?" may be either a research question having some theoretical significance or it may be a practical question, the answer to which will help to make decisions about funding, curriculum, etc.

It is important at the outset to determine what the purpose of the question is. This paper has contended that the simple question "Is Treatment A superior to Treatment B," is not sophisticated enough for those engaged in educational research designed to produce broad, generally applicable "laws."

Specific consumer use.

However, it may be sufficiently sophisticated for the consumer to ask when choosing between two specific instructional items. If the consumer is asking, is Textbook A better than Textbook B, given a particular set of students and a particular set of instructional objectives to be measured by an established evaluation system, then the question is appropriate. But the consumer must be forewarned first that the results and conclusions of his study will be as valid as the rigor with which the research has been conducted and secondly that the answer will be an extremely specific one. He will learn in a properly carried out study which textbook is better according to his definition of "better." He will not be justified in concluding any thing generally about texts as a method.

For the individual professor interested in improving his own course, it probably is not worthwhile to become involved in either complex and sophisticated research in education nor to be overly concerned with the method and treatment questions. The best strategy might be to specify his own course goals and construct some means of determining how well his students are progressing toward those goals. Then he might choose any innovative method which has some degree of legitimacy and with which he feels comfortable. Using broad guidelines for developing more effective instruction[6] can produce closer and closer approximations as he continues to make changes and observe the effects of those changes. This smacks of tinkering and it lacks elegance; the professor may not be able to contribute to the (insignificant) literature involving the comparison of treatments but he may end up with a much improved course.

The search ought to continue for generalizations which go far beyond the specific question: Should I use Text A or B? This is primarily a task for educational researchers who are likely to re-word the comparison question as stated at the outset of this article. Instead of comparing "methods" described in terms of formal properties, the contemporary researcher will try to isolate critical functions in teaching methods (e.g., feedback to students, structure or content) and study variations in these. Such work is going on apace in Psychology and Education.

Decision-making.

If research cannot tell us, à la Madison avenue, whether Brand A is better than Brand B, how can we make a decision when faced with such questions as: Should the University (the community, district, province) invest its resources in a particular innovation or teaching method? That question is being asked with increasing frequency today throughout the educational system. This article should have indicated that there is no easy answer. To obtain an answer for a specific comparison in a designated milieu carefully controlled studies can be mounted using similar populations, aimed at similar instructional goals and exposed to two differing treatments.

It is not merely the lack of data base that makes our original question unanswerable. Until recently we have lacked an adequately sophisticated engineering "science." In the gap between research and application--between the laboratory and the classroom--there is growing up an area of applied study which is likely to develop more and better means by which we can better define our questions and find ways of answering them.

References

1 Anderson, R. C. "Learning in Discussions: A resumé of the Authoritarian--Democratic studies." Harvard Educational Review (1959) 29, 201-215.

2 Dubin, R. Y. and Taveggia, T. C. The Teaching-Learning Paradox: A Comparative Analysis of College Teaching Methods. University of Oregon, Centre for Advanced Study of Educational Administration, 1968.

See also:

McKeachie, W. J. "Research on Teaching at the College and University Levels" in Gagne, N. L. ed. Handbook of Research on Teaching. Chicago: Rand McNally, 1963, (Chapter 23).

McKeachie, W. J. Research on College Teaching: A Review, Washington, D.C.: ERIC Clearinghouse on Higher Education, November, 1970, ED 043 789.

3 Anderson, Richard C. and Faust, Gerald W. Educational Psychology: The Science of Instruction and Learning New York: Dodd, Mead and Company, 1974.

4 Hohn, Robert L. "Effectiveness of Innovations in the Teaching of Psychology: A Critique" American Psychological Association Newsletter, December, 1973, 3-6.

5 Gagné, Robert The Conditions of Learning, 2nd Ed., New York: Holt, Rinehart and Winston, 1970.

6 See for example: Module #2: Characteristics of Effective Instruction in CLD's series: Teaching and Learning: An Individualized Course in Teaching Higher Education.